**Міністерство освіти і науки України  
 Національний технічний університет України  
 "Київський політехнічний інститут імені Ігоря Сікорського"  
 Фізико-технічний інститут**

Симетрична криптографія  
 Комп’ютерний практикум №1  
 Експериментальна оцінка ентропії на символ джерела відкритого тексту

Виконав:  
 студент групи ФІ-93  
 Папишев Олег

Перевірив:  
 Чорний О.М.

Київ 2022

ЗАГАЛЬНІ ВІДОМОСТІ

1.Мета роботи

Засвоєння понять ентропії на символ джерела та його надлишковості, вивчення та порівняння різних моделей джерела відкритого тексту для наближеного визначення ентропії, набуття практичних навичок щодо оцінки ентропії на символ джерела.

2.Постановка задачі

Написати програму для експериментальної оцінки ентропії на символ джерела відкритого тексту, порівняти різні моделі джерела відкритого тексту для наближеного визначення ентропії. Оцінити надлишковість російської мови у різних моделях відкритого тексту.

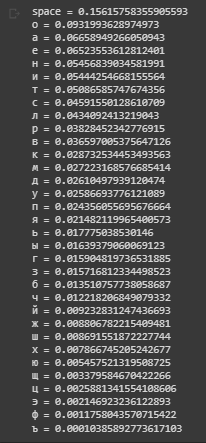
3.Хід роботи

1. Уважно прочитати методичні вказівки до виконання комп’ютерного практикуму.   
 2. Написати програми для підрахунку частот букв і частот біграм в тексті, а також підрахунку H1та H2 за безпосереднім означенням. Підрахувати частоти букв та біграм, а також значення H1та H2 на довільно обраному тексті російською мовою достатньої довжини (щонайменше 1Мб), де імовірності замінити відповідними частотами. Також одержати значення H1та H2 на тому ж тексті, в якому вилучено всі пробіли.

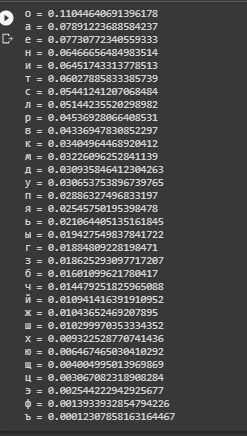
3. За допомогою програми CoolPinkProgram оцінити значення H(10) , H(20) , H(30) . 4. Використовуючи отримані значення ентропії, оцінити надлишковість російської мови в різних моделях джерела.

ПРАКТИЧНА ЧАСТИНА

1.Частота з пробілом:

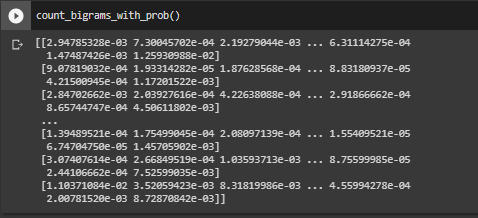


Частота без пробілу:



2. Частоти біграм:

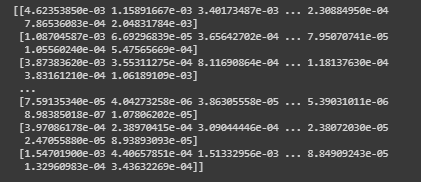
2.1) З пробілом:



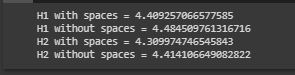
Нажаль гугл-колаб не хоче виводити матрицю 33х33, а виводити посимвольно я вважаю недоречним. Але за потреби можна легко отримати будь який елемент матриці.

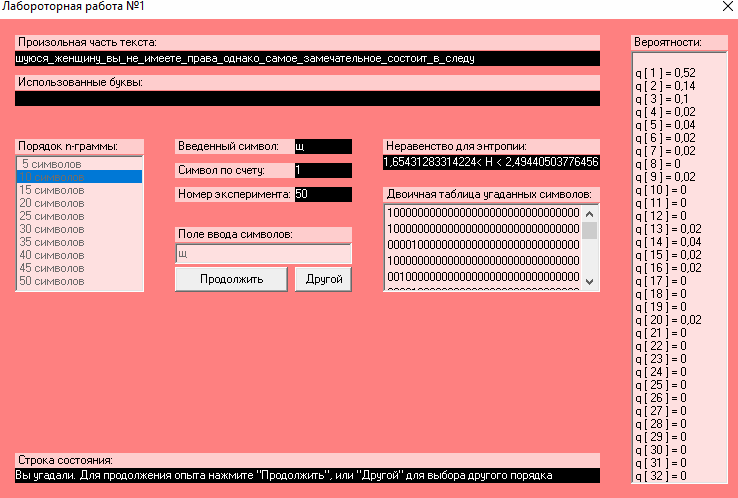
У стовпцях та строках перші 32 символи – літери руської абетки, без «ё», а 33 символ – пробіл. Біграми рахувались з перетином

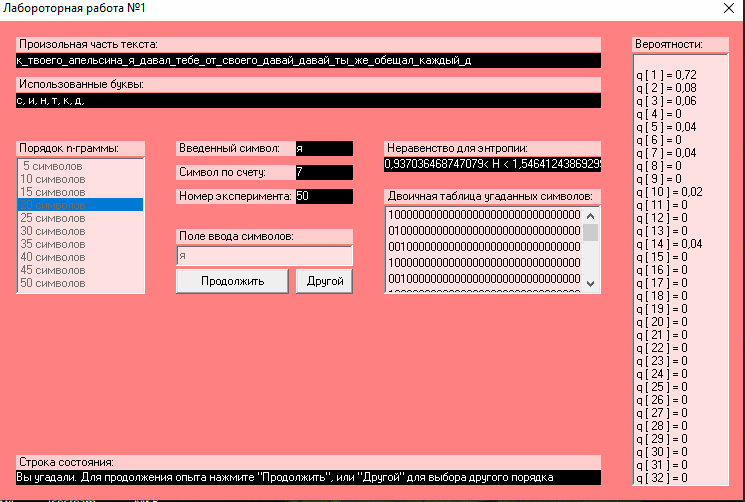
2.2) Без пробілу

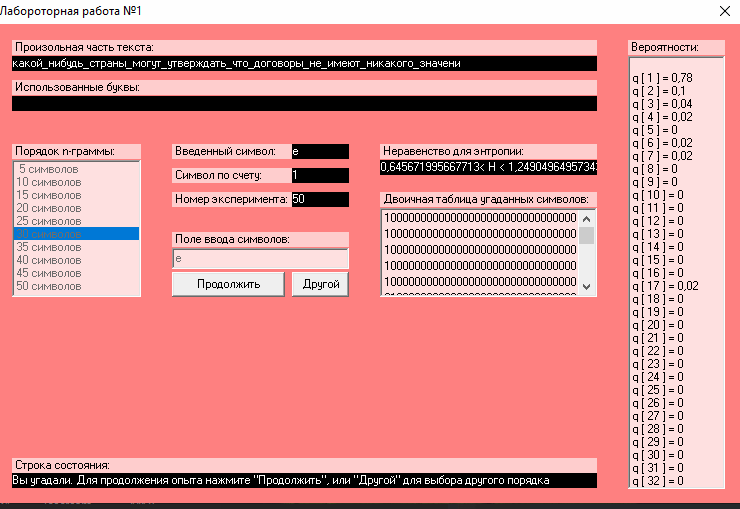


3.Ентропії:









4.Надлишковість російської мови:



H0 = log2(33)=5.044   
H\_inf =H30 = 0.645

R = 1 – 0.645/5.044 = 0.872

4.Висновки

Засвоїв понять ентропії на символ джерела та його надлишковості, вивчив та порівняняв різні моделі джерел відкритого тексту для наближеного визначення ентропії, набув практичних навичок щодо оцінки ентропії на символ джерела.